PROBLEMS OF DISEASE CLASSIFICATION IN MACHINE PROCESSABLE FORMAT

Thomas L. Lincoln M. D.

February 1968

P-3799

PROBLEMS OF DISEASE CLASSIFICATION IN MACHINE PROCESSABLE FORMAT

Thomas L. Lincoln M. D.*

The RAND Corporation, Santa Monica, California

Last May, Jesse Shera wrote a paper in Science on Information Technology entitled: "Librarians Against Machines."[1] He subtitled it with the remark that "Librarians are having difficulty adapting to the new technology because they have no professional philosophy." There is a parallel medical dilemma with respect to modern information handling which can be rephrased as follows: "Medicine is having difficulty adapting to new technology because medical and health personnel have so many professional philosophies."

It seems quite clear that this dilemma must be faced and resolved. Our professional philosophies are in part outmoded and in part outscaled by the present demand for information. It is comforting to recognize that this situation is a general one and not unique to medicine. The fact that this is a general situation may be useful in more closely defining the problem, and in pointing toward a solution. Then, the effective management of medical and health data will come to depend on a visible process of problem definition and therapeutic decision-making. This will take some reblocking and resorting of information, and ultimately, we can expect to change the ways we categorize disease and the elements we choose for clinical thinking.

Figure 1 illustrates what has happened to the diagnostic categories as we presently find them in textbooks. They are being pulled apart by at least three separate kinds of medical interests: the clinical, which looks at and treats each patient one at a time; the scientific, which

---

*Any views expressed in this Paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

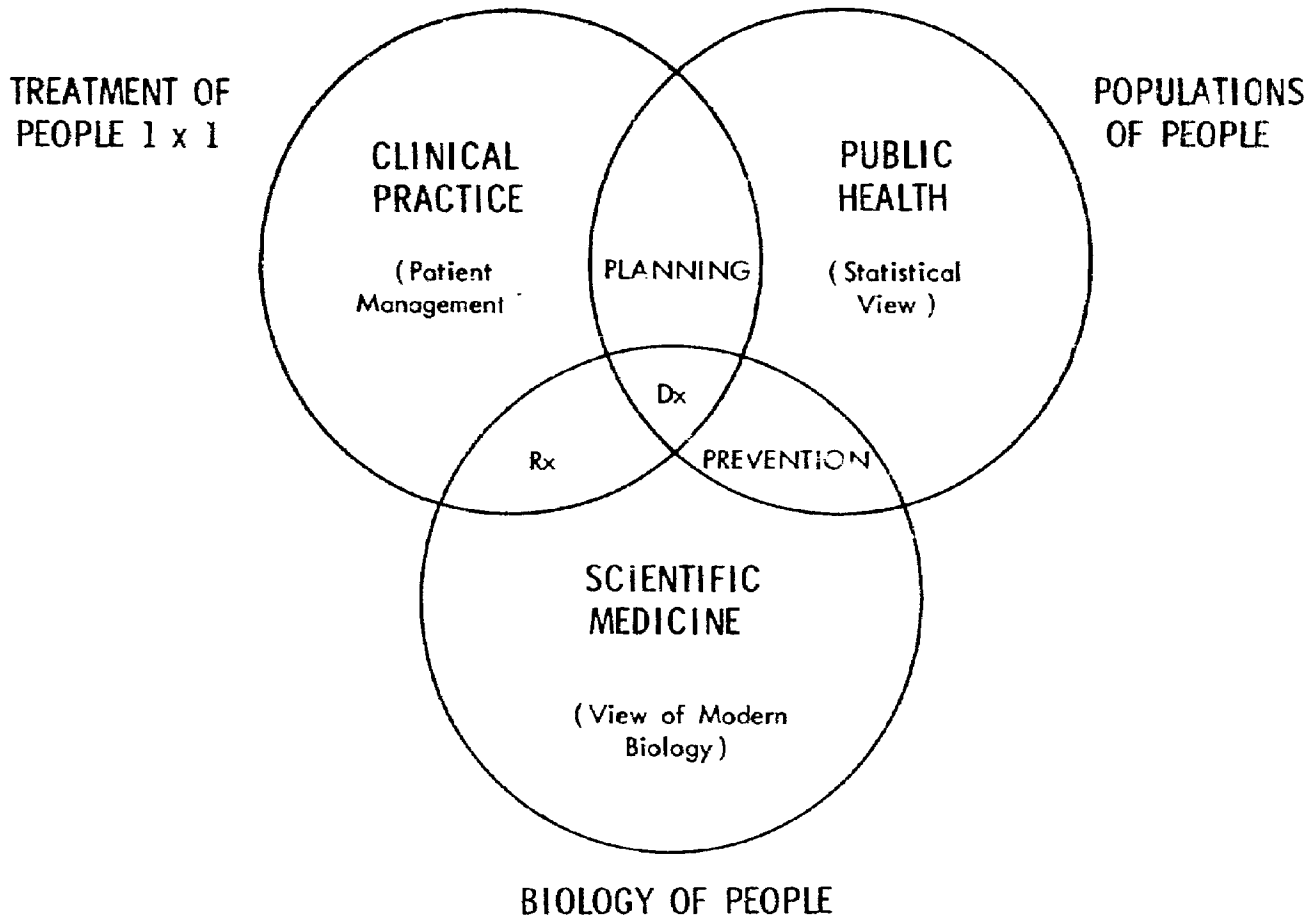

TREATMENT OF PEOPLE 1 x 1
CLINICAL PRACTICE
(Patient Management
PLANNING
PUBLIC HEALTH
(Statistical View)
POPULATIONS OF PEOPLE
Dx
Rx
PREVENTION
SCIENTIFIC MEDICINE
(View of Modern Biology)
BIOLOGY OF PEOPLE

Fig. 1

looks to uncover the pathophysiologic mechanisms of disease; and the epidemiological, which considers the effect of medical and health efforts on populations of people. These are district contexts, and use very different methodologies.

The clinical context requires information which is focused on the individual. It must address his problems and account for his particular constraints. The management of this situation is not difficult when a patient presents with a single correctable disease; however, this is not common, particularly in older people who require the most medical care. Here, the physician is faced with multiple problems which are never truly resolved. The management of multiple problems over a long period of time requires a maintenance routine of no small proportions.

The scientific context deals with the development of appropriate models which define disease processes. This is quite different from the clinical context because it concentrates on the principle rather than the people. The failure of a biological system and the failure of engineering systems offer striking parallels. At present we describe biological failure in terms of diseases and arrive at a particular disease model by a process of diagnosis. Engineering failure is described in terms of equipment defects and the process for finding the defect is termed troubleshooting. In addition to the parallels which can be drawn, there are important distinctions: the equipment has been constructed out of known components so that a defect is characterizable. In a biological model we can define a process to some level of distinctness, but then as we dig deeper, most medical problems become increasingly obscure. We have not built the system out of known parts and we do not really understand how it works.

The orientation of the epidemiological context is toward populations of people, and the evaluations are statistical. One purpose is to assess the behavior of a disease by looking at many examples. A purpose is to change the probabilities of disease by managing against the environment, by preventive measures such as early detection, or by changing personal habits of large numbers of people.

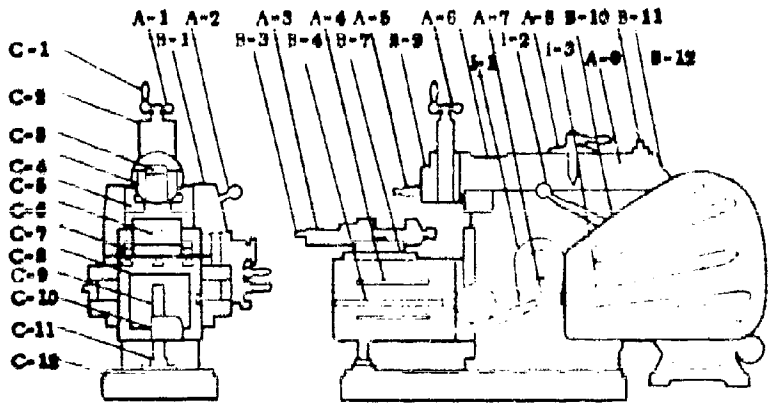
The interesting thing about these three contexts is that they utilize a common vocabulary of diseases. Many of our present and

potential difficulties in the management of medical information stem from the separate goals and methodologies of these overlapping languages. When there was less information, which could be handled more informally, this difficulty was less apparent. Likewise, the librarian faced with less information, could find informal ways of doing what seemed necessary without asking difficult operational questions. However, information processing technology requires precision, and the informal ways of fitting things together no longer work.

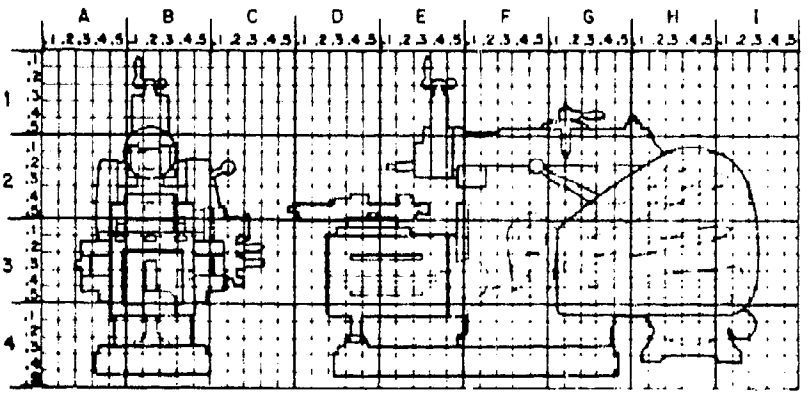
It may be instructive to look at this difficulty in a similar, but somewhat simpler context. A particularly insightful discussion of a parallel problem in electronic troubleshooting is presented in the Integrated Maintenance Manual by TECH/OPS, of Arlington, Virginia.[2] Figure 2 is taken from this manual and represents a problem which arises when a perfectly adequate notation for mechanical parts identification is scaled up. As more and more arrows are drawn in the diagram, it gets harder and harder to scan. The only return to economy is a change of notation. The grid at the right represents such a change. This is the simplest kind of an example. TECH/OPS provides several more which involve troubleshooting in radar transponders. Electronic circuit diagrams have suffered an extreme increase in complexity with the development of modern equipment. For a long time, it was possible for those familiar with the field to compensate for this by virtuosity so that persons adept at reading diagrams and at troubleshooting could stay ahead of the growing complexity. However, it becomes more economical to regroup the diagram into new functional elements and to consider operational components and component failures in a more structured way. Checkout algorithms can be written which focus on the defects.

I would like to suggest that the same is now demanded of medicine. The closest parallel is in the "wiring diagram" of biochemistry--the graph of relations among biochemical compounds which have been worked out over the past quarter century. The biochemical charts on laboratory walls have become more and more complicated so that the insightful pathways of an earlier time have become obscured by detail.

The increasing elegance and consequent obscurity of these charts is symptomatic of the present state of medicine. In a like manner,



600-A-1	TOOLHEAD SLIDE CONTROL, No. 00215-D5-4	10-45
600-A-2	TOOLHEAD SLIDE GEAR, No. 00265-M-7	10-45
600-A-3	TOOL SLIDE LOCK, No. 00987-Q-20	10-45
600-A-4	ELEVATOR CONTROL COVER, No. 00954-A-°	
[illegible]	TABLE SUPPORT CLAMP, No. 0087-BC-6	10-45
600-C-12	BASE, No. 00900-KC-97	10-45



REF DESIG	LOCATING COORD	NAME AND DESCRIPTION
600-A-1	B, 1/1, 1	TOOLHEAD, Giant Tool Mfg., Corp, Part No. 00215-D5-4
600-A-2	C, 2/2, 2	TOOLHEAD, Giant Tool Mfg., Corp, Part No. 00265-M-7
600-A-3	D, 2/2, 3	SLIDE LOCK, Giant Tool MFG., Corp, Part No. 00987-Q-20
600-A-4	D, 3/2, 3	ELEVATOR CONTROL COVER, Parts Co., Part No. [illegible]
[illegible]	[illegible]	[illegible] SELECTOR, Giant Tool Mfg. Corp, Part No. 00091-Z-8
600-I-2	F, 3/3, 3	TRAVERSE CONTROL, ABC Tool Co., Part No. 000987-A-8

Fig. 2

disease categories have been overrun by the available information and the diverse ways we have of collecting and using it. To give an example, diabetes is now an array of diseases with multiple causes and many levels of effects. Appropriate management is complicated by new drugs, new therapeutic regimens, and a change in the cases classified as diabetic. New screening methods turn up deviations from normal which are smaller than those previously considered and which are of unproven significance. Nevertheless they are often labeled as diabetes. There is a strong tendency to act on the label rather than on the individual and his problem.

Of those who have come to grips with this difficulty, I would like to cite Lawrence Weed,[3] Alvan Feinstein,[4] and Ralph Engle.[5] All three return us to a common thread of good sense: the fundamental medical unit is the patient, not the disease. Thus, we should "look in the bed before we look in the book."[3] The way out of our difficulty is to concentrate on elements which show stability in the face of rapid change and which reference the patient, and directly describe his problems. These elements include the sequence of acts which the physician undertakes, to define each problem, and to treat it. If the trace which generates the data is carried as an essential modifier of the data itself, then the context is not lost. This implies a problem-oriented medical record which is a record of patient maintenance, not just a collection of lists and orders. If these kinds of elements are used, then we can protect ourselves from the adverse consequences that arise in the manipulation of elements out of context. If patient problems are taken in the microparameters, it should be possible to cohesively relate data which results from the separate activities, interests, and goals of health professionals.

To do this, an appropriate sorting rule to define a patient problem must be stipulated. Unlike troubleshooting in engineering, the element cannot be "the defect." As noted above, we have never assembled a complex biological system and we cannot fully map its component parts. Thus, as we focus on a patient's problem, it is possible to select a set of questions which bring the problem under tighter definition, but this sequence gives out at some level. If we ask more questions, the

problem becomes vaguer and vaguer. In a logical sense, we are dealing with a partially convergent function, and the trick is to stop where there is maximum definition. Figure 3 attempts a simple diagram of this comparison.

For example, in hypertension, it is useful to rule in or out a pheochromocytoma. But, if indeed, the problem is an adrenalin producing tumor, it is not helpful to seek to more closely define the problem by asking about the cause of such tumors. We manage against it by taking it out.

The sorting rule becomes, then, the continuation of a diagnostic tree search to the point where it is possible to manage against the problem most effectively. The best description of a deviation becomes the point of maximum definition given the available information at a moment in time. The best definition carries the trace of the search in the appropriate stepwise fashion. This trace is dictated by the available information that is missing. Thus, a modestly elevated blood glucose in an asymptomatic person seen in a screening clinic will be noted to be just that. The test methodology, immediate past dietary history, and other relevant parameters, should be noted before a further search is instituted or a label appended to this deviation. If this is done, our separate medical interests will not mislead us. Given present data processing capability, it is possible to carry such relational information.

In summary, the interactions between points of view which define therapy, prevention, and medical planning (Figure 1) will achieve explicit unity only when this reformulation is done. If we are to achieve any serious orchestration of results in an increasingly complex health world, we will need to implement such systems.

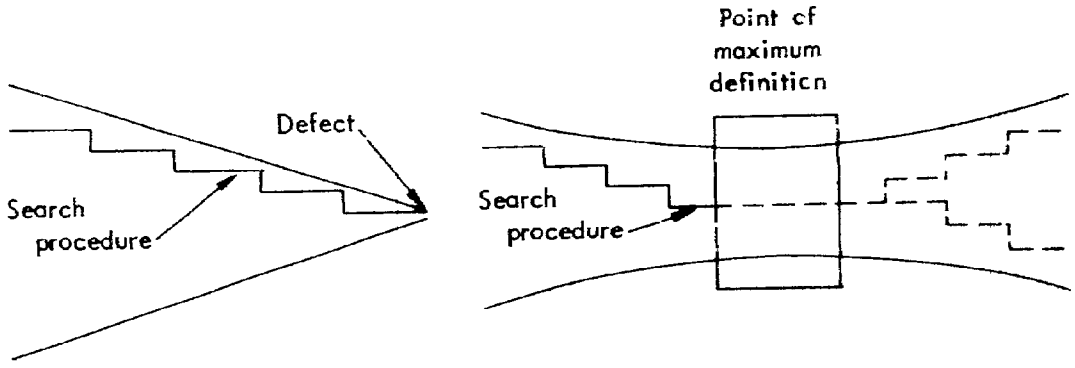

Point of
maximum
definition
Defect
Search
procedure
Search
procedure

Fig. 3

REFERENCES

1. Shera, Jesse H., "Librarians Against Machines," Science, Vol. 156, 12 May 1967, pp. 746-750.

2. The Integrated Maintenance Manual, TECH/OPS System Effectiveness Laboratory System Sciences Division, Technical Operations, Inc., Arlington, Virginia 22209. I am indebted to Adam Stein III, Electronics Division, United States Navy, for bringing my attention to this parallel and for several insightful discussions on the matter.

3. Weed, Lawrence, L., Problems That Confront the Physician and Some Possible Solutions (to be published).

4. Feinstein, Alvan, "Symptoms as an Index of Biological Behavior and Prognosis in Human Cancer," Nature, Vol. 209, No. 5020, January 15, 1966, pp. 241-245.

5. Engle, Ralph L., Jr., and Davis, B. J., "Medical Diagnosis: Present, Past, and Future," Archives of Internal Medicine, Vol. 112, October 1963.